Monitoring and Managing Hadoop with Ambari Web UI

Monitoring and managing Hadoop with Ambari web UI.
In this video, we will help your understand Apache Ambari.
And how you can use it to manage and monitor HDInsight, and
manage host and users through the Ambari web UI.
We'll also provide an overview of it's available services.
The Apache Ambari project is aimed at making Hadoop
management simpler by developing software for provisioning,
managing and monitoring Apache Hadoop clusters.
Ambari provides an intuitive easy-to-use Hadoop management
web UI backed by its RESTful APIs.
Ambari enables systems administrators to provision,
manage and monitor a Hadoop cluster.
Apache Ambari simplifies the management and monitoring of
a Hadoop cluster by providing an easy to use web UI and REST API.
Ambari is included on Linux based HDInsight clusters.
And as use to monitor the cluster to make configuration
changes.
Ambari makes Hadoop management simpler by providing
a consistent secure platform for operational control.
Ambari provides in it the web UI as well as a robust REST API,
which is particularly useful for
automating cluster of operations.
With Ambari, Hadoop operators get the following core benefits,
simplified installation, configuration and management.
Easily and effectively create, manage, and
monitor clusters at scale, takes the guesswork out of
configuration with smart configs and cluster recommendations.
Then your get centralized security set up,
reduce the complexity to administer and
configure cluster security across the entire platform.
Helps automate this set up and configuration of advanced
cluster security capabilities such as Kerberos and
Apache Ranger, full visibility into cluster health.
Ensure your cluster is healthy and
available with a holistic approach to monitoring.
Configures predefined alerts based on operational
best practices.
For cluster monitoring, captures and visualizes critical
operational metrics for analysis and troubleshooting.
It's also integrated with Hortonworks SmartSense, for
proactive issue prevention and resolution.
Highly extensible and customizable, enables Hadoop to
fit seemlessly into your enterprise environment.
Highly extensible with Ambari stacks, for
bringing custom services under management, and
with Ambari views for customizing the Ambari web UI.
The Ambari Architecture is based on the following design
principles.

Design goals, platform independence,
pluggable components, version management and
upgrade, extensibility, failure recovery, security, error trace
near realtime and immediate feedback for operations.
The Ambari web UI is available on your HDInsight cluster at
https://CLUSTERNAME.azurehdinsi-
ght.net where CLUSTERNAME is the name of your cluster.
Connecting to Ambari on HDInsight requires a HTTPS.
When connecting to the Ambari web UI,
you will be prompted to authenticate to the page.
Use the cluster admin user and
password you specified during cluster creation.
The metrics tab of the dashboard contains a series of widgets
that make it easy to monitor the status of your
cluster at a glance.
Several widgets such as CPU usage,
provide additional information when clicked.
The services side bar on the dashboard provides quick insight
into the status of the services running on the cluster.
Various icons are used to indicate the status or
actions that should be taken, such as a yellow recycle symbol
if a service needs to be recycled.
Cluster wide metrics show you information about your
whole cluster.
You can remove or add metric items from each cluster wide
metric widget, by selecting the item on the widget legend.
Get a larger view of the metric by selecting the magnifying
glass icon.
The Ambari dashboard includes these cluster wide metrics,
memory usage, the cluster wide memory utilization
including memory cashed swapped used in shared, network usage.
The cluster wide network utilization including in
and out.
CPU usage, cluster wide CPU information including system,
user and weight IO.
And cluster load, cluster wide load information including total
number of nodes, total number of CPUs,
number of running processes and one minute load.
The KPI widgets in the dashboard page have drilled down
capabilities that allow you to see detailed information for
the selected KPI.
The Ambari web interface home page is a dashboard
of the operating status of your cluster.
The dashboard contains metrics widgets
that provide status information for each service in your cluster
as well as the status of your cluster.
By default, you can see metrics for HDFS, YARN, and HBase.
In addition to cluster wide matrix in the Ambari dashboard,
you can add and remove individual widgets.
And rearrange the dashboard by dragging and

dropping each widget to a new location in the dashboard.
Depending on the service,
you can use the widget in the dashboard to display matrix.
You can create your own widgets or remove some widgets.
The HeatMaps tab displays metrics as colored heat maps,
going from green to red.
It provides a color coded view of each of the nodes in
the cluster for selected metrics.
Hovering over each of the nodes pops up additional information.
Cluster HeatMaps graphically represent how you used your
overall cluster with color coded blocks.
Each block represents a host in your cluster.
You can hover over a particular block to
see more information about that host.
A pop-up window displays metrics about Apache Hadoop components
that are installed on that host.
The colors that are displayed in the block represent usage
in a unit that is appropriate for the selected set of metrics.
Service config version history is available from the Ambari web
in two places.
On the Dashboard page under the Config History tab and
on its Service page under the Configs tab.
The Dashboard Config History tab shows a list of all version
across services with each version number and
the date and time the version was created.
You can also see which user authored the change
with the notes entered during save.
Using this table, you can filter, sort, and
search across versions.
Additional Config History details can be
seen by drilling down into the specific services.
This can also be seen from the Services view.
Ambari predefines the set of alerts that monitor the cluster
components and hosts.
Each alert is defined by an alert definition,
which specifies the alert type, check interval and thresholds.
Alerts other than okay will cause the number of Alerts Entry
at the top of the page to display the number of alerts.
Selecting this entry will display the alerts and
their status.
Alerts are organized into several different groups,
which can be viewed from the Alerts page.
Ambari provides many alerts which will have one of
the following as the status, OK, Warning, Critical, Unknown.
The alerts have drilled down capabilities allowing you to get
the end up information of the selected alert.
You can mange the alert groups by using the actions menu and
selecting Manage Alert Groups.
This allows you to modify existing groups or
create new groups.

This slide shows how to manage services using Ambari web UI.
The Services side bar on the dashboard provides quick insight into the status of the services running on the cluster.
Various icons are used to indicate the status or actions that should be taken such as yellow recycle symbol, if a service needs to be recycled.
Selecting a service will display more detail information about that service.
The actions menu button is where you would start, stop and restart an entire cluster.
You can also install additional services using this button.
For example, Falcon is not installed on this cluster and could be added using the Actions menu button.
Maintenance mode should be enabled when making cluster hardware or software changes.
For example, Maintenance Mode is useful when changing a service configuration, shutting down a cluster node, removing a cluster node or even during troubleshooting sessions.
Explicitly turning on Maintenance Mode for a service implicitly turns on Maintenance Mode for components or hosts that run that service.
Maintenance Mode suppresses Ambari alerts, warnings and status change indicators.
It also exempts an object from host level or service level bulk operations.
For example, if an HDFS data node eight were shut down for maintenance.
Ambari would generate alerts and warning messages about the data nodes failure.
Enabling maintenance mode on the data node prior to shutting it down would prevent this behavior.
As another example, if data nodes 11 and 12 are placed in Maintenance Mode.
Then they would be exempt from the bulk restart all or restart data nodes actions on the Ambari HDFS service actions menu button.
While Maintainance Mode prevents bulk operations, an Ambari user may still use the Ambari to explicitly start and stop a specific service, service component or host in Maintainance Mode.
From the Ambari web UI, select YARN from the list on the left of the page to see and view YARN service information.
You can see on the screen that shows the summary of YARN services.
And metric shows the resource civilization graph will perspective that last 1 to 12 hours and so on.
When the YARN service information is displayed, you can select the list of action that can be taken with YARN.
The Hive view, allows you to run Hive queries directly from your web browser.

You can save queries, view results, save results to
the cluster storage or download results to your local system.
Use Ambari hosts to manage multiple HDP components,
such as data nodes, name nodes, node managers, and
region servers running on hosts throughout your cluster.
To manage components running on a specific host,
choose a fully qualified domain name on the hosts page.
Summary components list all components installed on that
host.
Choose Options in Host Actions to start,
stop, restart, delete or turn on Maintenance Mode for
all components installed on the selected host.
On the Hosts page, you can drill down into the details about any
of the components running on the node.
This shows key metrics about the HDFS component
running on this node.
You can configure various HDFS parameters from here.
The YARN Capacity Scheduler allows Hadoop to be shared
among multiple independent tenants while
providing guaranteed capacity and predictable SLAs.
The Capacity Scheduler divides resources through use of YARN
queues which are sized based on the relative allocations given
to various tenants.
The Capacity Scheduler View solves these by
providing a simple UI that lets you create and
modify YARN queues and see their distribution at a glance.
The UI enforces configuration rules, highlights invalid
conditions and hides the complex syntax of setting ACLs.
The Tez View gives you visibility into all the jobs on
your cluster,
allowing to quickly identify which jobs consume
the most resources and which are the best candidates to optimize.
With the Tez View, you can quickly spot Hive or
Pig jobs that are taking the longest,
writing the most data or consuming the most CPU.
Once you've identified these big jobs, the Tez View lets
you drill in to see exactly how the job is running and
helps you identify ways to optimize it.
This view lets you see exactly how the job is executed and
the resources it uses at every step of the way.